

INRAE



Plan Données pour la science 2022-2024



Pour une transformation des pratiques et de la **culture numérique scientifique** à INRAE

Le document d'orientation stratégique [INRAE2030](#) souligne la nécessité de considérer les systèmes agricoles, alimentaires, environnementaux ou territoriaux dans leur globalité, leur diversité et leur complexité. Les avancées scientifiques et méthodologiques permises par le développement du numérique ouvrent de nouvelles voies pour explorer cette complexité, en intégrant et exploitant de vastes ensembles de données et de connaissances, en s'appuyant sur des approches *in silico* de modélisation, de simulation et d'analyses de données. L'accès aux données nécessaires pour ces approches est soutenu par la politique INRAE en termes de [science ouverte](#) et d'[infrastructure de recherche](#). Enfin, ces approches doivent s'inscrire dans des [usages responsables et sobres](#) du numérique.

Ce document présente la stratégie d'INRAE pour assurer la transformation de ses pratiques et de ses usages du numérique de recherche, ainsi que l'évolution des modes d'organisation associés. Ce véritable changement de culture qu'il est indispensable d'aborder de manière systémique, traverse quatre dimensions complémentaires : la science, l'ingénierie, les compétences et le partenariat. Les évolutions dans le domaine du numérique étant rapides, l'enjeu est d'impulser une dynamique collective agile et adaptable tant dans son contenu que dans ses modalités de mise en œuvre pour répondre aux enjeux. Il définit ainsi six objectifs stratégiques.



Introduction

Répondre à l'ambition de la stratégie INRAE2030 pour faire face aux grands défis en agriculture, alimentation et environnement nécessite des évolutions de nos usages et pratiques liés aux données, au sein de chaque communauté, à l'échelle de l'Institut et avec nos partenaires. Pour rendre notre recherche capable d'appréhender des enjeux disciplinaires, interdisciplinaires et transdisciplinaires en exploitant de manière approfondie et systématique le capital des données, notamment par des approches intégratives et prédictives, il est nécessaire d'inscrire cette transformation de nos pratiques dans le cadre d'une science plus ouverte. Cette évolution systémique concerne toutes les activités du cycle de la donnée et l'ensemble des communautés scientifiques et d'appui à la recherche.

Pour s'y préparer, plusieurs actions d'ampleurs diverses ont été initiées dans différentes communautés scientifiques, mais aussi au sein de l'appui à la recherche (e.g., Direction pour la Science Ouverte (DipSO) Direction des Systèmes d'information (DSI), nomination d'un administrateur des données, des algorithmes et des codes (ADAC), d'un chargé de mission numérique et d'une déléguée informatique et liberté). **Il apparaît essentiel de définir et partager une trajectoire d'évolution coordonnée permettant de conduire les projets transformants nécessaires. Il s'agit de le faire en synergie, pour que la donnée de recherche¹ soit un atout majeur de l'établissement pour l'ensemble de ses missions de recherche, d'innovation et d'expertise et appui aux politiques publiques.**

1. Conformément aux recommandations européennes et françaises, la donnée est considérée dans une acception large, incluant les tableaux de variables, les images, les enregistrements vidéo ou audio, mais aussi les codes sources informatiques et les modèles, à l'origine ou manipulant ces données.

2

Transformer les usages et les pratiques de la donnée de recherche et adapter l'organisation

Les usages, les pratiques et les modes d'organisation autour de la donnée de recherche varient d'une communauté scientifique à l'autre, que ce soit dans les étapes de son cycle de vie ou dans le niveau de maîtrise de la science des données. Dans ce contexte, il s'agit d'amplifier la dynamique sur les points suivant :

- Les données produites et gérées selon les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) sont considérées comme un produit majeur des recherches menées à INRAE. Partagés largement, les données et leurs usages sont un gage de l'intégrité requise pour des résultats reconnus par la communauté scientifique et la société ;
- La pratique de la réutilisation des données produites par des tiers doivent conduire à ne produire de nouvelles données que lorsque cela est nécessaire, gage d'efficacité et source de collaborations ;
- Les données et les activités associées constituent une source de connaissance, de reconnaissance et d'échange pour la recherche, l'innovation, et l'expertise et l'appui aux politiques publiques.

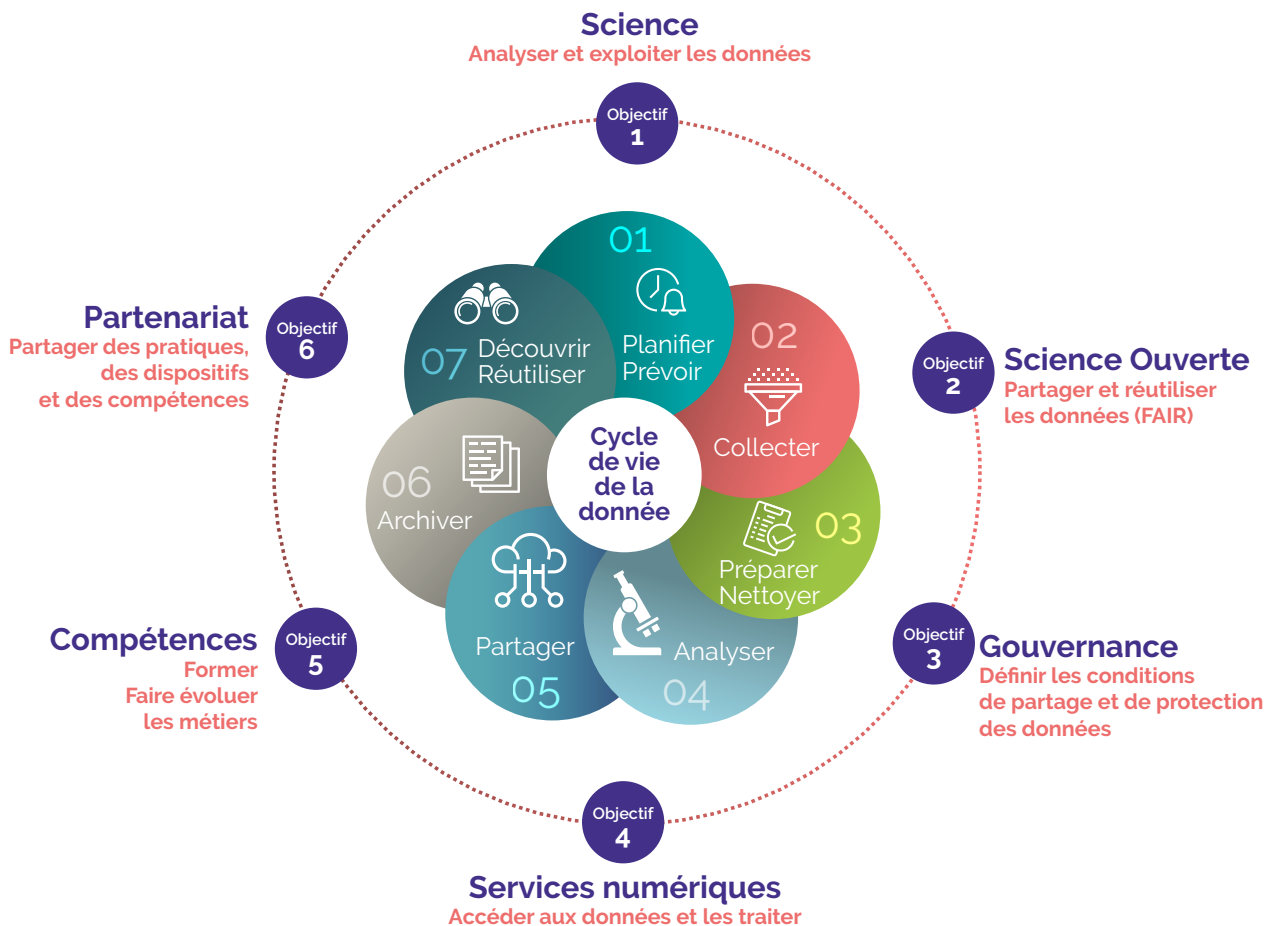
Une meilleure articulation est recherchée entre les activités d'acquisition de données, d'analyse de données et de simulation, au sein mais aussi entre communautés de recherches ou de l'appui. Dans ce cadre, les organisations de type e-infrastructure sont des instruments majeurs. Elles mettent à disposition des scientifiques des environnements informatiques de calcul et de stockage, ainsi que les appuis fonctionnels permettant une gestion, une exploration et une analyse des données efficaces et collaboratives. Il s'agit de trouver l'équilibre entre recherche de solutions génériques mutualisées d'une part, et adaptation aux besoins et questions spécifiques des communautés scientifiques d'autre part.

Cette transformation repose sur une implication de l'ensemble des agents et des organisations. Pour cela, il est nécessaire d'accompagner ces changements de culture, d'accroître les compétences en science et gestion des données, et d'adapter notre organisation et notre stratégie partenariale pour accompagner les différentes communautés.

Le plan données d'INRAE 2022-2024

6
objectifs

pour une transformation systémique
vers une science plus ouverte,
interdisciplinaire et prédictive



3

Six objectifs stratégiques pour les données

Cette politique de transformation des usages des données de recherche s'articule selon un nombre limité « d'Objectifs Stratégiques Données » (OSD). Ces objectifs permettent de prioriser et de coordonner les actions nécessaires pour leur déclinaison dans la stratégie des différents acteurs INRAE.

OSD1.

Accroître l'appropriation des nouvelles approches
en science des données par nos communautés scientifiques

L'extraction de connaissances nouvelles et originales à partir de données massives et hétérogènes est un enjeu majeur pour nos recherches disciplinaires et interdisciplinaires. Pour cela, il est nécessaire de développer l'usage des méthodes les plus récentes et les plus pertinentes de fouille, d'analyse et d'extraction de connaissances à partir des données, en développant nos capacités en traitement du signal, en statistique inférentielle, en modélisation, et en intelligence artificielle. Au-delà des méthodes de traitement des données sensu stricto, il s'agit d'encourager des changements en profondeur sur la place des données dans la démarche scientifique, depuis le choix des données à collecter jusqu'à leur exploitation (intégration dans des chaînes de calcul, assimilation de données, expérimentation numérique, etc.). Il s'agit ainsi de favoriser la mise en cohérence des activités de collecte, d'analyse et d'intégration des données d'une part, de modélisation et de simulation des systèmes étudiés d'autre part, pour répondre de façon optimale à des enjeux de compréhension, de prédiction et d'anticipation, de gestion et de contrôle, ou encore de conception. La mise en synergie de ces différentes activités, permettra de mieux comprendre, prédire, gérer et concevoir les systèmes complexes et dynamiques, et également d'identifier de nouvelles questions scientifiques.

OSD2.

Produire, partager et réutiliser des données
pour une recherche ouverte et reproductible

Un nombre croissant de nos projets de recherche mobilisent des données hétérogènes, multi-sources, stockées de façon distribuée. Dans ce contexte, la réutilisation de données doit être facilitée. Pour cela les données (incluant codes, workflows, modèles, etc.) doivent être suffisamment documentées et contextualisées pour permettre leur découverte et leur réutilisation. La mise en œuvre des bonnes pratiques de gestion de données s'appuie sur la rédaction de plans de gestion de données et sur l'anticipation de l'application des principes FAIR pour le partage des données. Cette meilleure gestion des données tout au long de leur cycle de vie contribue à la reproductibilité des résultats et globalement à la qualité des travaux de recherche. Il s'agit donc de poursuivre le développement, la diffusion et l'adoption des outils et des services, en appui à cette démarche, notamment dans le cadre des e-infrastructures avec la mise en œuvre d'outils et services génériques (portail de données, forge logicielle, ...) ou spécifiques aux com-



munautés (plan de gestion de données, ontologies de domaines, entrepôts et systèmes d'information thématiques, environnement de recherche virtuel...).



E-infrastructure de recherche ?

Une e-infrastructure de recherche est un dispositif informatique et organisationnel qui soutient les activités des scientifiques dans tout le cycle de vie de la donnée de recherche. Elle gère à travers ses services le stockage et le partage des données, leurs traitements, la mise à disposition de méthodes d'analyse, de modélisation et de simulation et l'accès à des ressources de calcul. Elle propose également un accompagnement de projets scientifiques dans son domaine d'expertise.

La plupart des communautés scientifiques ont organisé leur production de données au travers d'Infrastructures de Recherche (IR), structurées à différentes échelles (plateforme locale, IR nationale et Européenne). Ces infrastructures gèrent leurs données et leurs services numériques à travers une composante plus ou moins développée que l'on peut qualifier d'e-infrastructure.

OSD3.

Mettre en place une gouvernance des données accélérant partage, réutilisation et valorisation des données en toute confiance

La gouvernance des données, algorithmes et des codes met en place une organisation qui permet d'atteindre l'objectif d'un partage vertueux qui accélère la recherche et l'innovation au bénéfice des sphères scientifiques, politiques et socio-économiques. Un accompagnement des communautés scientifiques pour le partage et la valorisation de leurs données ainsi que la réutilisation de celles produites par des tiers est nécessaire pour que celles-ci se réalisent en toute confiance. Cette gouvernance fixe un cadre sécurisé donnant des repères et les appuis fonctionnels nécessaires aux collectifs de recherche. Pour cela elle précise le rôle et les responsabilités de chacun des acteurs et indique les modalités d'ouverture en fonction des risques, des bénéfices, et des contraintes qui s'y rapportent. La réutilisation des données partagées par des tiers est un levier économique important, mais également déontologique et éthique. Atteindre cet objectif, c'est travailler sur la confiance à pouvoir réutiliser les données produites par des tiers. Ceci est conditionné par le suivi des bonnes pratiques de partage de l'ensemble de l'écosystème scientifique qu'il s'agit de promouvoir.

OSD4.

Disposer d'infrastructures et de services numériques performants, résilients et interconnectés, facilitant les activités du cycle de vie de la donnée et les approches collaboratives

Chaque scientifique d'INRAE doit avoir accès à un environnement de travail numérique permettant d'accéder et d'intégrer facilement des données de différentes provenances (produites dans son unité ou ailleurs), et d'avoir accès à des services de calcul adaptés. Pour cela il s'agit d'intégrer les différentes briques technologiques et fonctionnelles des services proposés par les acteurs de la Direction des Systèmes d'information (DSI), de la Direction pour la Science Ouverte (DipSO), des plateformes, et des IR, afin d'offrir un environnement de travail facilitant l'usage des différents composants techniques pour le traitement de la donnée, du stockage au calcul,

incluant les outils de sécurisation, de visualisation et de développement. Cette intégration est faite en prenant bien en compte la nécessaire coordination avec nos partenaires de l'enseignement supérieur et de la recherche, les actions lancées dans le cadre de l'European Open Science Cloud (EOSC) et la dynamique des infrastructures de recherche nationales et européennes.

OSD5.

Accompagner l'évolution des compétences et des métiers autour des nouveaux usages

Le développement des compétences est un élément clé pour soutenir cette transformation. Cela implique que les personnels scientifiques et d'appui se forment régulièrement grâce à des dispositifs permettant de toucher le plus grand nombre tout en répondant aux besoins spécifiques individuels. L'évolution des compétences doit aussi être envisagée dans un contexte de changement rapide, tant des technologies que des dispositifs de recherche qui s'inscrivent dans des stratégies de mutualisation et de fédération au niveau national, européen ou international. Des actions sont en cours depuis plusieurs années sur ces dimensions tant au sein des départements de recherche, que des collectifs spécialisés (réseaux scientifiques, plateformes, collectifs d'informaticiens). Il s'agit de rendre visible, au niveau de l'établissement, ces actions et les coordonner pour mieux accompagner les agents. Enfin, l'évaluation tant individuelle que collective se doit également de prendre en compte ces nouvelles activités et ces temps de formations.

OSD6.

Conforter une stratégie partenariale pour accompagner les mutations liées à la nouvelle place des données dans les processus de recherche

L'ouverture des données de l'Institut et ses propres compétences en science des données lui permettent de renforcer ses partenariats, trouver des synergies au sein des écosystèmes régionaux (universités et grandes écoles), et aux niveaux national et européen, et d'attirer de nouveaux partenaires du monde socio-économique. Les communautés nationales actives dans les domaines d'INRAE s'appuient sur des infrastructures de recherche (IR), auxquelles INRAE contribue activement. Notre objectif est de contribuer aux ressources et services ouverts et partagés au sein des communautés scientifiques et du numérique de l'ESR, de participer aux actions d'animation et de s'inscrire dans la gouvernance de ces organisations, notamment les infrastructures de recherche. INRAE développe donc une stratégie partenariale, permettant d'articuler ces différentes dimensions, pour offrir à tous ses chercheurs un environnement de travail sur les données performant, facilitant l'accès à des sources de données, des outils de traitements, des services d'analyse, et à des compétences en science des données chez nos partenaires. Cela implique notamment d'aligner nos e-infrastructures de recherche aux IR nationales.



Conclusion

INRAE se dote d'une stratégie pour transformer ses pratiques et de sa culture numérique scientifique au travers de six objectifs (OSD). Cette stratégie se décline, via une gouvernance et une animation dédiée, dans les feuilles de route et projets des différents acteurs de la recherche et de l'appui à la recherche, afin que la transformation des pratiques des données de chacun bénéficie des actions de tous. La publication de ces OSD a pour objectif de partager cet engagement d'INRAE avec tous ses agents et ses partenaires, pour faciliter les collaborations dans des projets nationaux et internationaux, ainsi que le quotidien dans ses UMR.

Sites pour le partage des données

- Site <https://recherche.data.gouv.fr/fr> : plateforme de partage des données inter-instituts lancé par le gouvernement en juillet 2022. Chaque institut y possède son espace de dépôt des données. Une partie de cette plateforme a été développée par INRAE (lire l'article : <https://www.inrae.fr/actualites/recherche-data-gouv-donnees-commun>)
- Site <https://entrepot.recherche.data.gouv.fr/dataverse/inrae> : espace institutionnel de dépôt et de partage des données d'INRAE.
- Site data.partage d'INRAE <https://datapartage.inrae.fr/> : site donnant des indications et une boîte à outil pour rendre les données F.A.I.R (Faciles à trouver, Accessibles, Interopérables et Réutilisables. En particulier
- Où déposer les données <https://datapartage.inrae.fr/Partager-Publier/Choisir-un-entrepot>
- «Principes pour la gouvernance des données, des codes et des algorithmes». <https://datapartage.inrae.fr/Gouvernance-des-donnees2/Principes-pour-la-gouvernance-des-donnees-des-codes-et-des-algorithmes>
- Ouverture des codes, algorithmes, et logiciels. <https://datapartage.inrae.fr/Gouvernance-des-donnees2/Ouverture-des-codes-algorithmes-et-logiciels>
- Valorisation des données, codes sources, algorithmes et logiciels. <https://datapartage.inrae.fr/Zoom-sur/Valorisation-des-logiciels>



Centre-siège Paris-Antony
147 rue de l'Université – 75338 Paris cedex 07
Tél. +33 1 (0)2 42 75 90 00

Rejoignez-nous sur:



inrae.fr

**Institut national de recherche pour
l'agriculture, l'alimentation et l'environnement**



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*

INRAE